Journal of Pharmacy and Pharmacology 11 (2023) 56-59
doi: 10.17265/2328-2150/2023.03.002

3D Structural Analyses of Bifunctional Protein MdtA for Target Site Identification

Muhammad Zeeshan Ahmad

Department of Molecular Biology, University of Okara, Okara, Pakistan

Abstract: Bifunctional Protein MdtA catalyzes the dehydrogenation of methylene-H4MPT. MdtA also catalyzes the reversible dehydrogenation of methylene-H4F with 20-fold lower catalytic efficiency. Multiple structure prediction approaches including comparative modeling, threading, and ab initio were utilized to predict the 3D structure of the selected protein followed by the validation of the predicted structures through Errat, Procheck, and mol probity. The predicted 3D structure of MdtA revealed 9 alpha helixes sheets having 98.92% overall quality factor. Interestingly, it was observed that only 1.1% residues were present in the outlier region while 97.9% in favored and allowed region with -8.96 prosA z-score value. The selected protein participates in glyoxylate and dicarboxylate metabolism. In conclusion, the structural insight analyses of MdtA may improve the reversible dehydrogenation of methylene-H4F leads to comparative molecular docking analyses.

Key words: 3D structural insights, MdtA, in silico, Homology modeling, Bioinformatics.

1. Introduction

Methanol and methane are converted to CO_2 by methanotrophic and methylotrophic bacteria, respectively, through the use of formaldehyde and formate. Tetrahydromethanopterin (H4MPT) and its structural counterpart, tetrahydrofolate (H4F), are used by these microbes as C1 carriers in their C1 metabolism. NADP-dependent Methylorubrum extorquens AM1 is an enzyme of the catabolic C1 pathway that catalyses the stereospecific hydride transfer from H4MPT to $NADP^+$. Improved genetic tools are needed to empirically test new theories, such as cre-lox-based allelic exchange systems, transposon mutagenesis, and compact, broad-host-range plasmids for cloning, expression, and promoter probing. Utilization of formaldehyde is one issue with central metabolism in methylotrophy that calls for a novel genetic technique. The C1 unit is hydrolyzed by the NAD(P)-dependent methylene-H4MPT dehydrogenases MtdA and MtdB, as well as the formyltransferase-hydrolase complex Fhc, to produce formate and free H4MPT. Assimilation through the serine cycle uses the formaldehyde that condenses with H4F as the C1 donor. The H4F pathway's enzymes are typically three- to four-times more abundant during growth on C1 substances and are found at high specific activity during heterotrophic growth. The lack of growth on methanol is not explained by this, as formyl-H4F can be produced from formate during the methylotrophy process via the FtfL reaction [1].

Comparative models have been used to identify putative active sites and binding pockets, size of ligands, and relative affinities of ligands. Structural genomics is the goal of crystallizing proteins, or protein domains, to provide templates for families of related sequences for which suitable templates are lacking. When no suitable template is available, de novo modeling methods (also called ab initio modeling) may be used, but the success rate is lower than that with comparative modeling. Comparative modeling is used to compare two proteins with similar secondary and tertiary structures, even when determined under comparable conditions. Overall

Corresponding author: Muhammad Zeeshan Ahmad, research field: Bioinformatics. Email: ahmadjutt5028@gmail.com.

differences in protein backbone structures are quantitated with the root mean square deviation of the positions of alpha carbons, or rmsd. A model can be considered 'accurate enough' or as 'accurate as you can get' when its rmsd is within the spread of deviations observed for experimental structures displaying a similar sequence identity level as the target and template sequences. The 3DCrunch project used the SWISS-MODEL routines to do comparative modeling on all sequences in the Swiss-Prot database for which appropriate templates exist [2].

The most important details in this text are that up to 0.5 Å rmsd of alpha carbons occurs in independent determinations of the same protein, and that a highly successful comparative model must have ≥ 60% sequence identity with the target for a success rate > 70%. Even at high sequence identities (60%-95%), as many as one in ten comparative models have an rmsd > 5 Å *vs.* the empirical structure. The importance of the sequence alignment is also important, as misplaced indels can cause residues to be misplaced in space, and careful inspection and adjustment by someone with specialized training may improve the quality of the alignment and hence, the comparative model [2].

Comparative ("homology") modeling approximates the 3D structure of a target protein for which only the sequence is available, provided an empirical 3D "template" structure is available with > 30% sequence identity. In 2001, about 20% of sequences (in Swiss-Prot/TrEMBL) have suitable templates for comparative modeling at least part of the sequence. Comparative models are useful to get a rough idea where the alpha carbons of key residues sit the folded protein, guide mutagenesis experiments, or hypotheses about structure-function relationships. However, they are unreliable in predicting the conformations of insertions or deletions, as well as the details of side chain positions. In 2002, a new automated comparative modeling server came on-line: ESyPred3D [3].

The genetic diversity of the human proteome is now well understood because to the increasing amount of human genome population sequences. Proteins with little genetic variation may be found, and this strategy can now be applied to find 3D features and structures that are particularly intolerant of genetic variation. We hypothesised that preferred functional areas of the protein correlate to 3D characteristics that are intolerant to change. We used over 140,000 unique sequencing data points and more than 8,500 protein structural models to investigate this subject. The correlation between structural predictions and experimental functional readouts supported the theory. We think that data resulting from human variation complements other structurally-level criteria and can help guide medication development. Consider the case when you wish to determine a target protein's 3D structure but it hasn't been determined experimentally by NMR or X-ray crystallography. You just have the order. Software that organises the backbone of your sequence exactly like this template can be used if an experimentally established 3D structure for a protein that is sufficiently similar to yours (50% or better sequence identity would be nice) is available. "Comparative modelling" or "homology modelling" is what this is. In areas where the sequence identity is high, it is, at There are three inputs required for a comparative modelling routine:

1. The "target sequence" of the protein with the unknown 3D structure;
2. A 3D template is chosen based on which sequences most closely resemble the desired sequence. The template's 3D structure is normally a published atomic coordinate "PDB" file from the Protein Data Bank, and it must be established using trustworthy empirical techniques like crystallography or NMR;
3. The target sequence and template sequence are aligned.

Initially, the backbone is set up exactly like the

template's using the comparative modelling technique. This entails that the secondary structure, phi and psi angles, and alpha carbon locations are all designed to match the template exactly, most, relatively accurate for the placements of alpha carbons in the 3D structure. It is incorrect for added loops that don't have a matching sequence in the solved structure as well as for side chain position data [4].

2. Methods for the Prediction of Protein Interactions

Ramon Aragues and his coworkers used four different methods for predictions of protein-protein interactions:

(i). Gene fusion, in which two proteins are predicted to interact if their corresponding genes are fused in another genome [5].

(ii). Phylogenetic profiles, where similarity in phylogenetic profiles is interpreted as indicating that two proteins must be present concurrently to perform a given function together [6].

(iii). Distant conservation of sequence patterns and structure relationships, in which structural similarities among domains of known interacting proteins and conservation of pairs of sequence patches involved in protein–protein interfaces are used to predict putative protein interaction pairs [7].

(iv). Structural interologs, in which interactions are transferred between proteins with the same structural domains [8].

3. Drug Development Based on Protein Structure

The object of drug design is to find or develop a mostly small drug molecule that tightly binds to the target protein, either moderating its function or competing with natural substrates of the protein. Such a drug can be best found on the basis of knowledge of the protein structure. If the spatial shape of the site of the protein to which the drug is supposed to bind is known, then docking methods can be applied to select suitable lead compounds that have the potential to be refined into drugs [9].

4. Docking

Docking is a method that predicts the preferred orientation of one molecule when bound to another to form a stable complex. Predicting the strength of association or binding affinity between two molecules can be done using knowledge of the preferred orientations. Docking is frequently used to predict the binding orientations of small molecules and drug candidates to protein targets, which in turn predicts the small molecule's affinity and activity. The development and implementation of a range of molecular docking algorithms based on different search methods were observed in the last few years [10].

5. Conclusions

Computational methods for protein structure prediction are still in the early stages of development, and methods like homology-based prediction become especially helpful in an environment where the methods can be used in concert with experimental techniques for the structure and function determination of proteins. The use of computers and computational methods permeates all aspects of drug discovery today and forms the core of structure-based drug design. In the modern drug discovery process, the availability of 3D protein structures, high-performance computing, data management software, and the internet facilitate access to a massive amount of generated data and transform the massive, complex biological data into workable knowledge. Computational tools offer the advantage of delivering new drug candidates more quickly and at a lower cost. Protein with two functions Methylene-H4MPT is dehydrogenated using MdtA as a catalyst. MdtA also catalyses the reversible dehydrogenation of

methylene-H4F with a 20-fold lower catalytic efficiency. The 3D structure of the chosen protein was predicted using a variety of structure prediction techniques, such as comparative modelling, threading, and ab initio, and then the predicted structures were validated using Errat, Procheck, and mol probity. MdtA's estimated 3D structure showed 9 alpha helix sheets with an overall quality factor of 98.92%. It was interesting to see that just 1.1% of residues were in the outlier area, compared to 97.9% in the preferred and permitted region with a prosA z-score value of -8.96. The chosen protein takes involvement in the metabolism of glyoxylate and dicarboxylate.

References

[1] Huang, G., Wagner, T., Demmer, U., et al. 2020. "The Hydride Transfer Process in NADP-dependent Methylene-tetrahydromethanopterin Dehydrogenase." *J Mol Biol* 432 (7): 2042-2054.

[2] Hicks, M., Bartha, I., di Iulio, J., et al. 2019. "Functional Characterization of 3D Protein Structures Informed by Human Genetic Diversity." *PNAS* 116 (18): 8960-5.

[3] Martz, E. 2001. "Comparative ("Homology") Modeling for Beginners with Free Software." Available at: https://www.umass.edu/microbio/chime/pe_beta/pe/protexpl/homolmod.htm.

[4] Bagchi, P., Mahesh, M., and Somashekhar, R. 2009. "Pharmaco-Informatics: Homology Modeling of the Target Protein (GP1, 2) for Ebola Hemorrhagic Fever and Predicting an Ayurvedic Remediation of the Disease." *Journal of Proteomics & Bioinformatics* 2 (7): 287-294.

[5] Enright, A. J., Iliopoulos, I., Kyrpides, N. C., & Ouzounis, C. A. 1999. "Protein Interactionmaps for Complete Genomes Based on Gene Fusion Events." *Nature* 402 (6757): 86-90.

[6] Aragüés, R., García-García, J., & Oliva, B. 2008. "Integration and Prediction of PPI Using Multiple Resources from Public Databases." *Journal of Proteomics & Bioinformatics* pp. 166-187.

[7] Pellegrini, M., Marcotte, E. M., Thompson, M. J., et al. 1999. "Assigning Protein Functions by Comparative Genome Analysis: Protein Phylogenetic Profiles." *Proc Natl Acad Sci U S A* 96 (8): 4285-8.

[8] Espadaler, J., Romero-Isart, O., Jackson, R. M., & Oliva, B. 2005. "Prediction of Protein-protein Interactions Using Distant Conservation of Sequence Patterns and Structure Relationships." *Bioinformatics* 21 (16): 3360-8.

[9] Ramanathan, K., Karthick, H., and Arun, N. 2010. "Structure Based Drug Designing for Diabetes Mellitus." *Journal of Proteomics & Bioinformatics* 3 (11): 310-313.

[10] Chikhi, A., and Bensegueni, A. 2008. "Comparative Study of the Efficiency of Three Protein-Ligand Docking Programs." *Journal of Proteomics & Bioinformatics* 1 (3): 161-165.